From Dataset to Melody: Enhancing Music Composition with Computational Models

Valton Kamberaj[1]

Arbana Kadriu[2]

Nuhi Besimi[2]

[1]PhD Student,
Faculty of Contemporary Sciences and Technologies,
"SEEU", University,
Tetovo, North Macedonia
[2]Prof. Dr.,
Faculty of Contemporary Sciences and Technologies,
"SEEU", University,
Tetovo, North Macedonia

Received: 23 May 2025 / Accepted: 17 July 2025 / Published: 01 August 2025

Doi: 10.56345/ijrdv12n2007

Abstract

This study explores the generation of melodic lines in the ethno-fusion genre using computational models, leveraging a unique MIDI-based dataset. The dataset, designed with two primary dimensions—solos and chords—aims to ensure precise, genre-specific outputs. The core techniques employed are Markov chains and generative grammar, chosen for their suitability in generating sequences aligned with rhythmic, pitch, and structural characteristics of the dataset. Generative grammar focuses on chord generation, isolating it from solos to enhance harmonic coherence, while Markov chains facilitate the creation of both solos and their combination with chords. The results, validated through extensive testing and feedback from music professionals, demonstrate high accuracy and utility in producing new compositional ideas, particularly within the Balkan music industry. This work provides a foundation for further exploration of AI-driven music generation, emphasizing scalability to other dimensions and instruments.

Keywords: *Music generation, MIDI dataset, Markov chains, Generative grammar, Ethno-fusion, Computational models*

1. Introduction

The intersection of music and computer science is an exciting and rapidly growing field, but it still faces many challenges and gaps (Martineau, 2008). Music, at its core, is an art form that uses a delicate interplay of frequencies to evoke emotions, tell stories, and share ideas. Beyond its role as entertainment, music also serves as a tool for therapy and social connection (P. Kasák et al., 2020; Huang & Yang, 2020). For researchers working to bring music and technology together, the quality of input data—whether it's sound, symbolic representations like sheet music, or audio signals—is critical. Without accurate data, AI systems can't effectively analyze or create music, as poor inputs limit their ability to learn and produce meaningful outcomes (Silla et al., 2016; Oore et al., 2018). This data often comes in formats like MIDI files or raw audio signals, where the combination of various frequencies forms the basis of musical compositions (Kasák et al., 2020; Huang & Yang, 2020).

At its essence, music is about organizing frequencies into melodies, harmonies, and rhythms, each of which shapes its emotional and structural impact. With advancements in AI, especially in machine learning and neural network models, there is growing potential for technology to create music that mirrors human creativity. These models treat music like a language, breaking it down into smaller, manageable pieces that a computer can process. For example, by turning a musical score into a sequence of time-based events, AI systems can compose expressive pieces, such as MIDI files that emulate the nuance of classical music (Choi et al., 2020; Oore et al., 2018; Payne, 2019). Modern innovations like sparse attention and recurrence mechanisms, used in advanced models such as Transformer-XL, have made it possible for machines to capture the long-term patterns in music, enabling them to generate more complex and sophisticated compositions (Yi-Jen Shih et al., 2023; Zou et al., 2022; Qin et al., 2023).

The fusion of computer science and music has already shown incredible potential to transform the music industry. Programming, algorithms, and machine learning are now essential tools for enhancing music creation, classification, and analysis. Music, often called a universal language, plays a powerful role in shaping social identity and bringing communities together. Technology is helping musicians and researchers refine their craft and explore new possibilities. However, as the music industry becomes more commercialized, some genres and compositions risk losing their value, making the tasks of classification and preservation all the more critical (Kong et al., 2022; Shiromani et al., 2023; Liao et al., 2022). The integration of computer science into music offers not only a way to preserve musical integrity and improve genre classification but also opens doors to entirely new forms of creative expression.

This paper elaborates on our joint work, which involves testing two algorithms, various techniques, and methods to obtain the results. The input for these algorithms and techniques is a dataset we created in MIDI format.

Our dataset, explained in the paper (Kamberaj et al., 2024), is based on two precisely created dimensions for each symbolic sequence. The first dimension represents the melodic line created by different notes and chords which are created by combining three different frequencies for each one. The third dimension which is not presented with any specific sound but is only created based on it is the tact which also represents the basic part without which the combination of notes and music in general would have no meaning at all (Kamberaj et al., 2023).

The basis of each built sequence is pitch duration, timbre, sequential length and timing accuracy (Kamberaj et al., 2024). In this paper, we emphasize/describe the importance of four dimensions, which we implemented into algorithmic ideas to generate consecutive and complete sequences of a melodic line. We specifically chose to focus on two dimensions—solos and chords—because, based on our review of related works, this approach yields greater accuracy than incorporating additional dimensions.

2. Literature Review

According to researchers, these frequencies can take two main forms that are particularly suitable for creating datasets and testing as inputs. Music, as an art form, is fundamentally a composition of sound frequencies designed to evoke emotions and convey ideas. Its structure relies on the careful arrangement of notes, rhythms, and harmonies, each playing a crucial role in shaping the overall aesthetic and emotional impact. In music theory, rhythm—or meter—provides the framework for organizing notes, with beats forming the foundation of musical structure. Without rhythm, music loses its coherence, becoming a disjointed collection of sounds (Hao-Wen Dong et al., 2018; Gatti et al., 2017). Beyond aesthetics, music serves as a therapeutic tool and fosters community connections by expressing cultural identities and shared values (Martineau, 2008).

The digital representation of music through MIDI (Musical Instrument Digital Interface) has transformed both music analysis and generation. MIDI sequences, which encode note events like pitch, timing, and velocity, have become a standard for studying music. They enable researchers to explore computational models for generating musical sequences (Gómez & Kroher, 2015). MIDI not only supports music analysis but also forms the backbone for training machine learning models to create new compositions (Shibata et al., 2021). For instance, datasets of piano-based MIDI sequences are widely used due to the piano's predictable frequency patterns, which make it ideal for generating accurate musical representations (DuBreuil, 2020; Kong et al, 2021). However, using MIDI data from instruments like guitars introduces challenges due to their more complex and variable frequency structures (Sarmento et al, 2023; Donahue et al., 2019).

One of the earliest methods for generating music involved Markov Chains, which model the probability of one note following another. These models are especially effective for generating short, coherent musical phrases. Higher-order Markov models, which consider longer sequences of notes, can create more intricate musical structures (Kong et al, 2022). However, their ability to maintain long-term coherence in music remains limited, pushing researchers toward more advanced techniques.

Another powerful approach is Generative Grammar, which uses predefined rules to construct musical sequences, similar to the grammar of natural languages. By applying these rules, music generated through this method adheres to established conventions of rhythm and harmony. This technique has shown promise in creating compositions that align with specific styles and rules of music theory (Yang & Wu, 2023; Wu et al., 2020.

In recent years, neural sequence models have revolutionized music generation. Techniques such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) are designed to process sequential data while retaining memory of past events. This allows them to generate more complex musical patterns (Xianchao Wu et al., 2020; Sarmento et al., 2023). Despite their successes, these models face limitations when it comes to capturing long-term dependencies, often struggling with coherence in extended compositions due to challenges like vanishing gradients (Donahue et al., 2019).

To address these limitations, Transformer models have emerged as the leading technology for music generation. Leveraging self-attention mechanisms, Transformers can process entire sequences simultaneously, making it easier to capture long-term dependencies. These models have proven exceptionally effective in generating music with expressive structures and intricate variations (Child et al., 2019; Payne, 2019). Innovations like Transformer-XL, which incorporates recurrence mechanisms, have further improved their ability to produce coherent and dynamic musical compositions (Oore et al., 2018). Additionally, Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) offer alternative approaches. VAEs learn compact representations of musical sequences, enabling the generation of new compositions by sampling from their latent spaces (Choi et al., 2020). GANs, with their generator-discriminator architecture, refine compositions through adversarial training, ensuring that generated music is both diverse and authentic (Silla et al., 2016).

The quality of datasets remains a critical factor in music generation. Custom-built MIDI datasets often outperform pre-existing datasets, especially when they closely align with the target genre or instrument (Kong et al., 2021; DuBreuil, 2020). The piano-roll approach, which visually represents musical sequences on a grid, has been particularly effective, offering precise control over pitch, timing, and dynamics (Qiu et al., 2023). However, generating music for other instruments, such as guitars, continues to present unique challenges due to their intricate frequency structures (Sarmento et al., 2023; Wu, 2020).

Despite remarkable progress, challenges remain. Ensuring the musicality and emotional resonance of AI-generated compositions is still an open question (Child et al., 2019). Additionally, achieving style transfer—where music is generated in a specific style or genre—requires a deep understanding of the elements that define different musical traditions, making this an ongoing area of research (Gatti et al., 2017). The road ahead promises exciting possibilities as researchers continue to refine these methods, pushing the boundaries of what is possible in AI-driven music generation.

The future of music generation lies in improving the expressiveness and creativity of AI-generated compositions. As generative models continue to evolve, they will enable more personalized and interactive music creation, where AI assists musicians and composers in generating novel ideas and compositions. AI-driven tools, in combination with human creativity, have the potential to transform the music industry by facilitating new forms of artistic collaboration and innovation.

3. Methodology

The fundamental elements of music that define the meaning of a melodic line are solos, chords, and rhythm, which together represent the synchronizable time-dependent relationships between these components. The chord establishes a harmonic foundation, providing a contextual framework for the solo, and thereby implying the overall meaning of the melodic line. A chord is formed by the combination of at least three distinct notes played simultaneously, serving as the harmonic basis upon which a single note or a sequence of notes in the solo part is played or anticipated. The solo, in turn, consists of individual notes played within the rhythmic structure, which must remain harmonically compatible with the chords defined in the corresponding time interval. The combination of many different frequencies one after the other represents the solo part. Both of which are of great importance and only with one of these is it possible to create a melodic line by producers or composers (Ferretti, 2017; Mauch et al., 2012; Gao et al., 2022).

Our work focuses on the generation of solos, chords, and their combination, culminating in the machine's ability to produce a two-dimensional melody. We employ two techniques-generative grammar and Markov chains- which are specifically applied to MIDI-based data. These algorithmic approaches generate musical sequences from our dataset, with the generative grammar method sequentially combining elements to produce chords that align with the genre and fit the required note combinations. Typically, our system generates melodic lines in combinations of three distinct beats, although the majority of MIDI melodies in our dataset adhere to a 4/4 time signature.

Using generative grammar, we separate the solo and chord components. Based on pitch and duration data from the dataset, the system generates three additional notes to form a new chord. This chord is constructed to align with the harmonic framework of the genre, enabling the generation of a compatible solo that completes a melodic line. The sequential length of the chord fits within the defined beat and is tailored to the dataset, with an average tempo of 120 bpm. Depending on the melodic structure, sequences may occupy 100% of a beat or conclude at 80–90% of it. If a chord extends beyond 50% of the beat duration, it can serve as the basis for generating solo sequences corresponding to each chord. This approach has consistently yielded favorable results, and extensive testing has demonstrated its effectiveness. Feedback from music industry professionals has highlighted its utility in creating new and innovative melodic lines.

Using the Markov chains technique, we implemented two distinct methods: one focused on generating the solo part and the other on generating solos combined with chords. When the system predicts the next sequence based on the dataset and generates only solos, it produces accurate outputs that align with the note transitions characteristic of the specified genre.

In the second method, where both solos and chords are generated together, the system achieves similarly strong results despite the increased complexity of working in two dimensions. These outcomes have been rigorously tested by music industry professionals, who have successfully used the generated results as a basis for creating new melodic lines.

The dataset we have created so far (Kamberaj et al., 2024), will be used to get the melodic lines that we have created.

The dataset is created in MIDI format for reasons of greater accuracy during the algorithmic work. Most of it is in the 4/4 time signature, then we have two other rhythms such as 2/4 and 7/8 which have a smaller part of the creation. We have also created some melodic lines in combination with other instruments for testing purposes to measure accuracy, but in this paper we present our generation with the melodic lines from our dataset created only with piano-roll.

4. Results

Using our dataset and the implemented algorithmic techniques, we have generated the results presented in this chapter. The results are categorized based on two primary algorithms: generative grammar and Markov chains.

4.1 Generative Grammar generation

Using the portion of the dataset that represent the majority of the data the results will be presented in visual form in 4/4 time, based on a tempo range of 110-135 bpm. Since this technique focuses solely on the chord part, separated from the solo part, the chords are formed by combining three sequential notes—representing the minimum foundation for generating chords with the highest accuracy using the smallest combination of notes. Separating the solo part allows the algorithm to perform a type of training in which it identifies a combination of three notes and then effectively generates the subsequent chord sequences. The combination of these three frequencies can be saved as MIDI output for playback and as a (.xml) file for a visual representation of the notes on a musical staff. This visualization can be achieved using a powerful tool like MuseScore. Below, we present a figure showing the output of one of the tests conducted on our dataset.

Figure 1: Generative grammar generation (MuseScore, presentation of results).
Source: Author

Referring to Figure 1, we note that the notes created and their combination belong very much to the folk and oriental genre, which also describes the logic of creating our dataset in that direction. Considering that the melodic lines are 97% created in A minor, it is easy to see the combination of the A, C and E sequences in one of the chords and that the future melodic lines regarding the solo can be freely built based on these chords in this scale, and why not then be modulated in other notes as well, since when it is created in one, it can then be "transposed" to the others. The verification of the accuracy of the sequence generation is done in two ways in our case. First, the chord generation can be heard by professionals and its accuracy can be determined, and then the presentation in MuseScore, which simultaneously gives notes on the pentagram as in the figure shown, and can also be played, seeing with accuracy which sequence is being executed, and then an assessment of what has been achieved to help the music industry.

4.2 Markov chains

The generation process this technique is divided into two parts. The first part focuses on the solo, as opposed to creating chords. In this part, each sequence within the 4/4 beat separates the solo from the chord part and generates new melodic line sequences. Extensive testing has been conducted to evaluate this method, yielding excellent results. Feedback from music professionals highlights its effectiveness in creating new melodic lines, while the subsequent placement of chords can be easily managed by producers. The core aspect of this technique is the simultaneous generation of complete chords and solos to construct a full melodic line. In our testing, we successfully generated sequences in the "Arabic" time signature, which occasionally employs 4/4 for the bass and, in certain cases, 2/4—effectively the same, except for two fewer beats, while remaining within the time signature limits for this rhythm. From the generation phase, we present a melodic line that demonstrates our work and the combinations of notes derived entirely from the custom dataset. This dataset is unique and the first of its kind in this genre, offering potential for use by producers across the Balkan region.

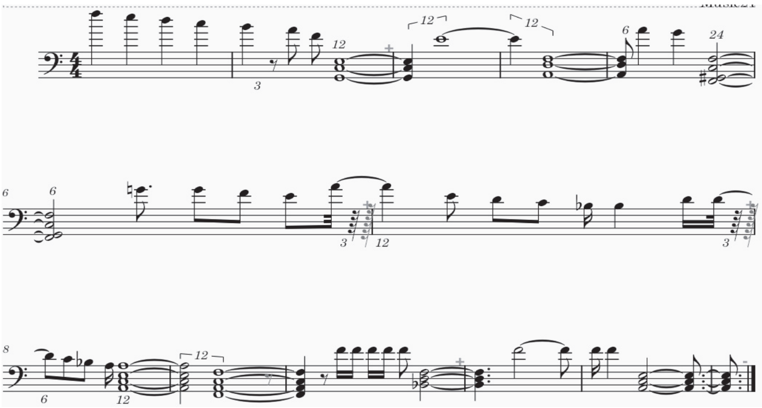

Figure 2: Markov chains generation (MuseScore, presentation of results).
Source: Author

From the higher notes generated as separate sequences in the solo and chord parts, we notice that the note that is most commonly used in our genre is touched, which if played in A minor is B minor.

4.3 Other visual analyses and comparisons

From the generation tests, we are also visually presenting some of the main characteristics of the fusion of a melodic line from the two techniques, which are: rhythm, pitch, frequency, distribution and finally accuracy, which will be compared to one of the inputs with our dataset in 4/4 rhythm.

4.4 Note Pitches Histogram

The histogram below (Figure 3) we can analyze the number of notes and their frequencies used in generating the melody line.

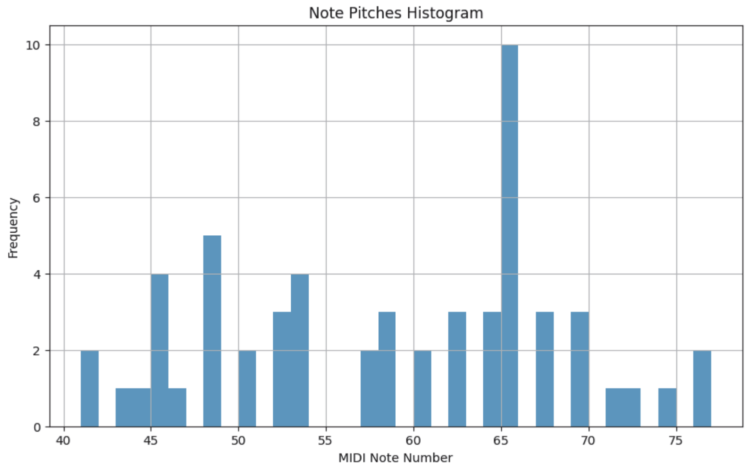

Figure 3: Note pitch histogram (generated melodic line).
Source: Author

The notes that are used according to the ordinal numbers in the wide list of notes in octaves that start from C-1 and continue up to G9. If we see the range of notes approximately from 40 - 80, then we notice that the entire melody line is within these numbers which are then characterized by the frequency that presents each note at which height of the scale it is played.

The generated melody is in the A minor scale, with notes falling within the same range and closely matching one of the lines from our dataset. This alignment will be explained in greater detail later. Referring to the standardization of musical notes, ranging from the lowest to the highest frequency, we can observe a strong alignment of the generated frequencies with the corresponding notes. This alignment demonstrates that the notes are played within the same octave. This consistency serves as clear evidence that the generation process has been successful in adhering to the trained parameters of octaves, scales, tempo, and beat, all within the limits of our dataset.

4.5 Note Time Events

In Figure 4 we will present the analysis always based on the melodic line that we have generated which includes note times events.

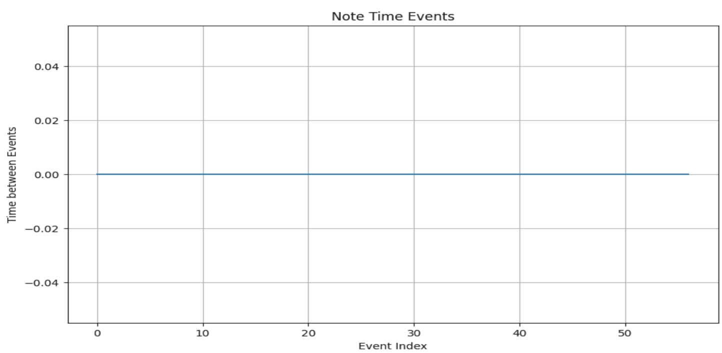

Figure 4: Note time events (generated melodic line).
Source: Author

If we refer to the Figure 4, we notice that in time between events is the blue line where until the end of the melody it has the value 0. In fact this value indicates the exact and 100% synchronization of all solo notes and chords which are created to start at the same time and end at exactly the same time taking into account the frequency synchronization then the 4/4 time and the complete accuracy of the parameters as far as this generation is concerned. The beginning and ending of the notes, mainly the foundation that is built with chords, begins exactly at the beginning of the measure and ends in the same way, and then the solo part also stays within the limits, synchronizing not only with the chords but also with the other frequencies of the solo within.

4.6 Rhythm Distribution of Notes

Figure 5 presents the rhythm distribution, describing the work of the basic dimension which is the rhythm for the generated line.

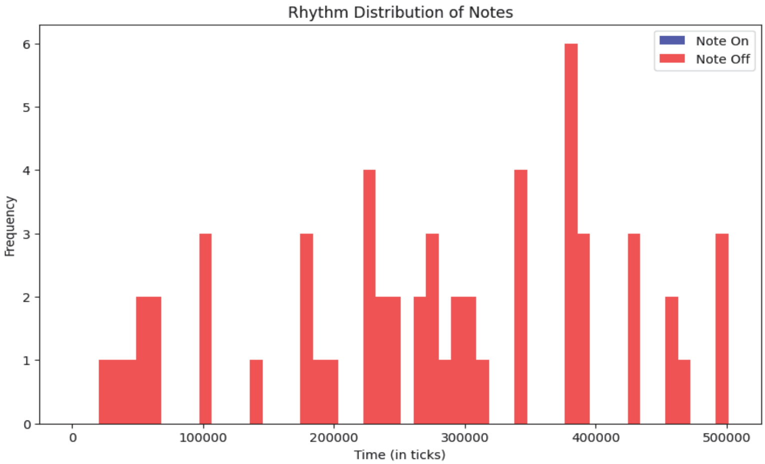

Figure 5: Rhythm distribution of notes (generated melodic line).
Source: Author

Note that the above representation is in time ticks and the corresponding frequencies on the X and Y axes. The results obtained from our first generation show that the most important dimension is present in which the notes are distributed sequentially and the strata end at their frequencies without leaving the main 4/4 beat. If we connect it to figure 4 where we present that notes like solo and chords do not have a time gap for some notes to start earlier and some later but then when it is necessary to be precise with the rhythm domain then each sequence of notes is distributed in proportion to the rhythm created that based on the tempo then and the frequencies that are created we can see notes that have a longer length and some a shorter length and also some more pronounced than the other depending on how sensitive the music is if it is more rhythmic with more pronounced frequencies or more melancholic that a softness in the notes is required. As for the chords part, the start and end of each of the 3 notes for a combination for a chord is not valid because they must start equally and end in the same way with the same length and with the same sensitivity. While the solo part differs as seen in the figure because not all notes have the same length and frequency or the same start. Here the most important thing is that the start and end of the solo part be in 4/4 time in this case.

4.7 Distribution of note durations

The next comparison is about the distribution of note durations, which is shown on Figure 6.

The note values shown in this figure represent the frequency of whole notes and half notes used within the melodic line. Considering that we are in A minor then the notes used are also combined with the same notes from the chord part and are expressed in frequency which shows the value of how many times such a note is used. It is clearly seen that each part of the note is created in its place and if we compare it with the representation in the pentagram we notice that each frequency has the exact number of touches or appearances in this melodic line to once again verify the generation.

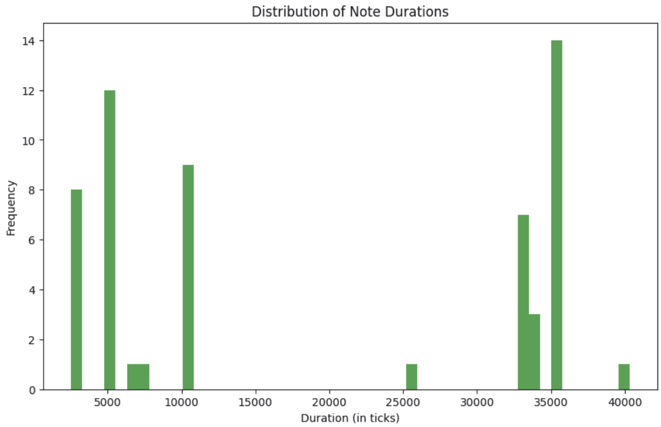

Figure 6: Distribution of note durations (generated melodic line).
Source: Author

4.8 Comparison with one of melodic lines (our dataset)

The comparison in this section belongs to one of the created melodic lines which is in 4/4 rhythm with a tempo of 125 bpm with the generated music which we are presenting in Figure 7.

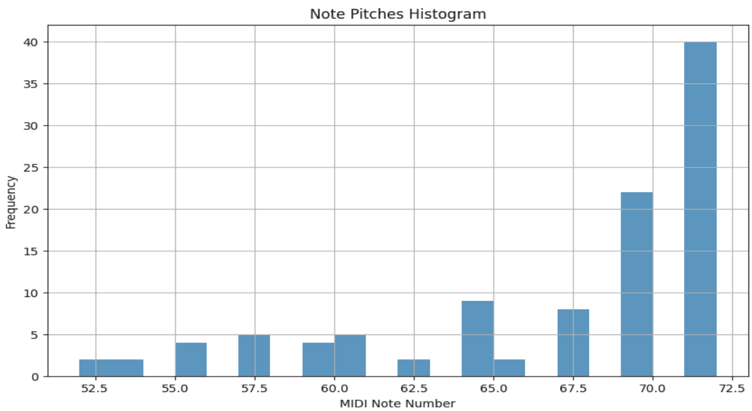

Figure 7: Note pitch histogram (melodic line from dataset).
Source: Author

The comparison is made with Figure 3 where it is observed that the number of notes on the X axis starts from 50 and goes up to 71, while in the aforementioned figure from 43 to 76. The accuracy of first the scale of playing the notes and then the octaves is clearly observed where these figures also show the minor scale played under A both in the generated music and in the line from the created dataset. In fact, the distribution of notes is approximately equal, and now on the Y axis the frequencies change slightly because the melodic line is longer and the notes may have repetitions because if the chorus is played 2 or 3 times, those repeated and identical notes indicate the number of frequencies in the music, while as for the generated result, it is shorter and the notes that are repeated have a smaller number of appearances, a difference that does not affect the accuracy in the most important dimensions such as the beat that has the "Arabic" rhythm and then in the notes within the frames of both the solo and the chords.

5. Conclusions

Generating music and providing new ideas through machine learning is highly challenging but also offers significant support to producers. The aim of this work is to generate music in the ethno-fusion genre using a unique dataset, contributing to the support of producers from Balkan countries by offering new compositional ideas. The primary techniques used are Markov chains and generative grammar, chosen to build the logic necessary to adapt to our dataset and achieve the most accurate generation. Considering the availability of various techniques, and referring to prior works in this field—some of which have yielded excellent results—we focused on methods that align with the main dimensions of music, specifically the two fundamental components: solos and chords.

The generation part based on the training of our dataset as regards generative grammar has given very good results based on the 4/4 rhythm and the combination of chords with 3 notes for each. Referring to these results and comparing them with some of the melodic lines of the dataset, the results including the different dimensions are very good based also on the responses of many producers who have seen it as a great help in creating new ideas.

The Markov chains technique includes numerous tests and numerous analyses from which in this paper we have presented only the best and most accurate results, showing that we have generation in the solo part alone, as well as in the combination with chords. As for the generation of the solo alone, we have had very good results with which even that part, according to the producers, is a very helpful part then to generate only the chords and to compose a melodic line. According to the figures and visualization of the results for the main dimensions for each generated sequence, we note that all elements are within the value frames with which we created the dataset, thus including the rhythm, the repeated notes with frequencies within the limits, and the intervals that occur at the limits of the beat and there is no time interval in which the solo or chord part goes outside it within 4/4. We conclude that the generation of values is very close to the created dataset and the generated lines are helping to create new ideas that will be used in the music industry by Balkan producers.

5.1 Limitations and future work

In this direction, there is still a lot of work to be done because the idea is to add more dimensions and to try with other instruments and not just the piano. The limitations so far are the dimensions and beats other than 4/4 which will be the target in other works once they are added to the dataset. Generation can also give great direction through the following techniques with which the accuracy will be further increased and the symbolic notes can be turned into audio by adding another instrument, or the parts can be separated so that the solo can be played with any instrument by the instrumentalists and then the chords can be generated or the chords can be played on the guitar or piano and the solo part can be generated later.

References

Chris Donahue, H. H. (2019). LakhNES: Improving multi-instrumental music generation with cross-domain pre-training. arXiv (pp. 1-8). Ismir: arXiv. Retrieved from https://doi.org/10.48550/arXiv.1907.04868

DuBreuil, A. (2020). Music Generation with Magenta. Birmingham: Packt.

Ferretti, S. (2017). On the modeling of musical solos as complex networks. Elsevier, 271-295. Retrieved from https://doi.org/10.1016/j.ins.2016.10.007

Gómez, N. K. (2015). Automatic Transcription of Flamenco Singing From Polyphonic Music Recordings. IEEE/ACM Transactions on Audio, Speech, and Language Processing,, 901-913. doi:10.1109/TASLP.2016.2531284

Hao-Wen Dong, W.-Y. H.-C.-H. (2018). MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). (pp. 34-41). Vancouver: The Thirty-Second AAAI Conference. doi:10.1609/aaai.v32i1.11312

Kristy Choi, C. H. (2020). Encoding Musical Style with Transformer Autoencoders. Proceedings of the 37th International Conference on Machine Learning (pp. 1899-1908). Viena: PMLR. Retrieved from https://proceedings.mlr.press/v119/choi20b.html

Lorenzo Gatti, G. Ö. (2017). Automatic Generation of Lyrics Parodies. Proceedings of the 25th ACM International Conference on Multimedia (pp. 485–491). New York: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3123266.3123410

Martineau, J. (2008). The Elements of Music: Melody. Rhythm, and Harmony. Bllombsbury Publishing, 1-6. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Martineau%2C+J.+%282008%29.+The+Elements+of+Music%3A+Melody%2C+Rythm+and+Harmony.+Bllombsbury+Publishing+&btnG=

Matthias Mauch, H. F. (2012). Integrating Additional Chord Information Into HMM-Based Lyrics-to-Audio Alignment. IEEE Transactions on Audio, Speech, and Language Processing, 200-210. doi:10.1109/TASL.2011.2159595

P. Kasák, R. J. (2020). Music information retrieval for educational purposes - an overview. International Conference on Emerging eLearning Technologies and Applications (ICETA) (pp. 296-304). Košice: IEEE. doi:10.1109/ICETA51985.2020.9379216

Payne, C. (2019, May 13). MuseNet.com. Retrieved from towardsdatascience: https://towardsdatascience.com/openais-christine-payne-on-ai-generated-music-and-the-road-to-super-intelligence-580f5063093

Q. Kong, B. L. (2021). High-Resolution Piano Transcription With Pedals by Regressing Onset and Offset Times. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 3707-3717. doi:10.1109/TASLP.2021.3121991

Qiuqiang Kong, B. L. (2022). GiantMIDI-Piano: A large-scale MIDI dataset for classical piano music. arXiv, 1-11. Retrieved from https://doi.org/10.48550/arXiv.2010.07061

Rewon Child, S. G. (2019). Generating Long Sequences with Sparse Transformers. arXiv, 1-10. Retrieved from https://doi.org/ 10.48550/arXiv.1904.10509

Ruchir Shiromani, T. M. (2023). Analysis of Automated Music Generation Systems Using RNN Generators. (eds) Decision Intelligence Solutions (pp. 237–247). Springer. Retrieved from https://doi.org/10.1007/978-981-99-5994-5_22

Sageev Oore, I. S. (2018). This time with feeling: learning expressive musical performance. Springer Nature Link, 955–967. Retrieved from https://doi.org/10.1007/s00521-018-3758-9

Sarmento, P. K. (2023). GTR-CTRL: Instrument and Genre Conditioning for Guitar-Focused Music Generation with Transformers. Artificial Intelligence in Music, Sound, Art and Design (pp. 260-275). Switzerland: Springer Nature. Retrieved from https://doi.org/10.1007/978-3-031-29956-8_17

Silla, C. N., Przybysz, A. L., & Leal, W. V. (2016). Music education meets computer science and engineering education. IEEE Frontiers in Education Conference (FIE) (pp. 1-7). PA, USA: Erie. doi:10.1109/FIE.2016.7757413

Valton Kamberaj, A. K. (2023). Instruments Music Composition in Different Genres and Techniques Using AI: A Review. (eds) Economic Recovery, Consolidation, and Sustainable Growth (pp. 741–754). Springer Nature. Retrieved from https://doi.org/10.1007/978-3-031-42511-0_48

Valton Kamberaj, A. K. (2024). COMPARATIVE SYMBOLIC ANALYSIS OF THE ETHNO-FUSION GENRE: INSIGHTS AND PERSPECTIVES . SEEU Review, 65-80. doi:10.2478/seeur-2024-0075

Valton Kamberaj, A. K. (2024). Creating the Benchmark Dataset for Ethno-Fusion Music: From Piano Rolls to an Ensemble of Instruments. 2024 13th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). Budva: IEEE. doi:10.1109/MECO62516.2024.10577816

Xianchao Wu, C. W. (2020). Transformer-XL Based Music Generation with Multiple Sequences of Time-valued Notes. arXiv, 1-9. Retrieved from https://doi.org/10.48550/arXiv.2007.07244

Xiaoxue Gao, C. G. (2022). Automatic Lyrics Transcription of Polyphonic Music With Lyrics-Chord Multi-Task Learning. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2280-2294. doi:10.1109/TASLP.2022.3190742

Yang Qin, H. X. (2023). Bar transformer: a hierarchical model for learning long-term structure and generating impressive pop music. Springer Nature, 10130–10148. Retrieved from https://doi.org/10.1007/s10489-022-04049-3

Yang, S. -L.-H. (2023). MuseMorphose: Full-Song and Fine-Grained Piano Music Style Transfer With One Transformer VAE. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 1953-1967. doi:10.1109/TASLP.2023.3270726

Yi-Jen Shih, S.-L. W.-H. (2023). Theme Transformer: Symbolic Music Generation With Theme-Conditioned Transformer. IEEE Transactions on Multimedia, 3495-3508. doi:10.1109/TMM.2022.3161851

Yikai Liao, W. Y. (2022). MICW: A Multi-Instrument Music Generation Model Based on the Improved Compound Word. (pp. 1-10). Taiwan: IEEE. doi:10.1109/ICMEW56448.2022.9859531

Yu-Siang Huang, Y.-H. Y. (2020). Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Compositions. MM '20: Proceedings of the 28th ACM International Conference on Multimedia (pp. 1180 - 1188). New York: ACM. Retrieved from https://doi.org/10.1145/3394171.3413671

Zipeng Qiu, R. Y. (2023). Generated Therapeutic Music Based on the ISO Principle. Communications in Computer and Information Science (pp. 32-45). Singapore: SOMI. Retrieved from https://doi.org/10.1007/978-981-97-0576-4_3

Zou, Y., Zou, P., Zhao, Y., Zhang, K., Zhang, R., & Wang, X. (2022). Melons: Generating Melody With Long-Term Structure Using Transformers And Structure Graph. ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 191-195). Singapore: IEEE. doi:10.1109/ICASSP43922.2022.9747802